



DATASHEET · CAPTURE · VOICE STT

Typing up meetings word for word? From now on, every recording comes back as text.

Voice STT turns recordings of meetings, calls and dictations into text. It detects the language on its own, and the Ultra tier separates the speakers. Built for office, sales and project teams, in the app and via API.

AT A GLANCE

Audio and video files up to 50 MB, automatic language detection, three quality tiers, speaker separation in Ultra, and word timestamps, SRT and VTT via API.

01 · WHAT IT IS ABOUT

First the meeting, then the typing? Now the text arrives by itself.

The project meeting minutes have to go out today, and the recording runs an hour. With Voice STT you upload the file and get the text back, separated by participant in Ultra, ready to copy or translate.

INPUT

Audio (MP3, WAV, OGG, FLAC, M4A) and video (MP4, WebM, AVI, MOV) up to 50 MB

LANGUAGES

Detected automatically; 100 language codes via API, 15 languages to pick in the app

OUTPUT

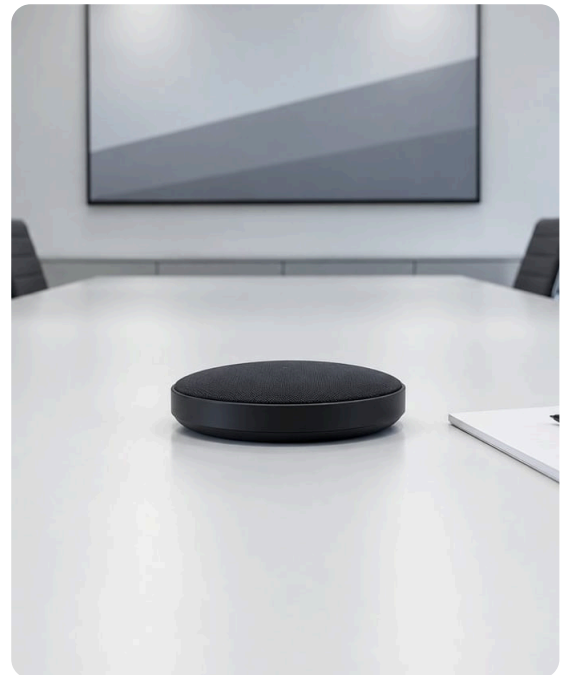
Transcript as plain text, in Ultra as a dialogue per speaker; download as TXT

ACCESS

App under

Media AI › Voice Studio

, REST API and MCP



Language detected automatically

Voice STT detects the language of the recording on its own. If you know it, you can set it, via API from 100 language codes.

Three quality tiers

Basic returns fast plain text, Premium adds segments with word timestamps, Ultra also separates the speakers. The tier is under [Transcription quality](#).

Separate the speakers

In Ultra, Voice STT detects the participants and shows the transcript as a dialogue. Rename speakers with one click.

Vocabulary

With [Vocabulary](#), Voice STT recognizes technical terms from your knowledge bases and price lists more accurately. Via API you add your own hotwords from Premium.

Subtitles via API

Via API, Premium and up returns SRT and VTT plus word timestamps, and Ultra adds RTTM for speaker analysis.

On to Translate Studio

[Use in Translate Studio](#) hands the transcript straight to Translate Studio, with no copying and pasting.

02 · HOW IT WORKS

Drop the file, pick a tier. Use the transcript.

In Voice Studio it takes four steps. Via REST API and MCP you send audio files straight from your own processes.

1

Open Voice Studio

Under `Media AI › Voice Studio`, choose the `Speech to text (STT)` tab. From the start page, `Quick Tools › Media › Speech to text` takes you there too.

2

Drop the file

Drag the audio or video file into the field. Leave the language on Auto or pick it.

3

Pick tier and options

Under `Transcription quality`, choose Basic, Premium or Ultra. In Ultra you can set the number of speakers under `Conversation partner detection`.

4

Use the transcript

Copy it, save it as a text file with `Download as TXT`, or hand it to Translate Studio with one click.

TECHNICAL DATA

FEATURE	DETAILS
Input formats	MP3, WAV, OGG, FLAC, M4A, MP4, WebM, AVI, MOV; in the app up to 50 MB, one file per run
Languages	automatic detection; 100 language codes via API, 15 fixed choices in the app
Quality tiers	Basic (API default), Premium, Ultra (app default)
Output via API	output_format json, srt and vtt from Premium, rttm in Ultra
Speaker separation	in Ultra, optional min_speakers and max_speakers from 1 to 20
REST API	POST /job/add/paperoffice_voice___stt with audio_file
MCP	po_stt_transcribe on /mcp-full

03 · IN PRACTICE

Meeting, dictation, training video. Each as finished text.

Minutes from a project meeting

The project lead uploads the recording in Ultra, renames Speaker 1 and 2 to their names and copies the dialogue into the minutes.

Dictation from the field

A surveyor dictates into a phone on the road and uploads the M4A file at the office. Basic returns the plain text for the report, and the vocabulary knows the product names.

Subtitles for a training video

Development sends the video via API with `quality=premium` and `output_format=srt` and gets the subtitle file with timestamps back.

FREQUENTLY ASKED QUESTIONS

Which languages does Voice STT recognize?

The language is detected automatically. Via API you can set it from 100 language codes, and the app offers 15 languages as fixed choices.

Can Voice STT tell speakers apart?

Yes, in Ultra. The transcript appears as a dialogue, and speakers can be renamed. Via API you also get the speaker count, overlaps and RTTM.

Can I get subtitles as SRT or VTT?

Yes, via API from the Premium tier with `output_format=srt` or `vtt`. In the app you save the transcript as a TXT file.

How is it billed?

By usage, depending on the quality tier. Current prices are on the pricing page.

GET STARTED

Let your next recording be written up for you.

Sign in to your PaperOffice account, open **Media AI › Voice Studio** and choose **Speech to text (STT)**. Drop the file, pick a tier, start the transcription.

paperoffice.ai/en/ai-voice-stt-speech-to-text-transcription-recognition-stt



MORE ABOUT THE MODULE

paperoffice.ai/en/ai-voice-stt-speech-to-text-transcription-recognition-stt

PRICING

paperoffice.ai/en/pricing

CONTACT

paperoffice.ai/en/contact